DOCUMENT RESUME

ED 070 799 UD 013 110

AUTHOR	Williams, Robert L.
TITLE	The BITCH-100: A Culture-Specific Test.
INSTITUTION	Washington Univ., St. Louis, Mo.
SPONS AGENCY	National Inst. of Mental Health (DHEW), Bethesda, Md.
PUB DATE	Sep 72
NOTE	21p.; paper presented at the American Psychological Association Annual Convention, Honolulu, Hawaii, September 1972
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Achievement Tests; Culture Free Tests; Dropouts; High School Students; Inner City; *Intelligence Tests; Negro Culture; Negro Dialects; Racial Differences; *Test Bias; *Test Construction; Testing Problems; *Test Validity
IDENTIFIERS	BITCH 100; Black Intelligence Scale of Cultural Homogeneity; California Achievement Test

ABSTRACT

The aim of this paper is to describe the rationale and evaluation of the Black Intelligence Scale of Cultural Homogeneity (BITCH). A "culture specific" test is used to determine the taker's ability to function symbolically or to think in terms of his own culture and environment. A combination of dialect specific and culture specific tests would certainly enhance the possibility of measuring what is inside the black child's head; this is the basic rationale for the BITCH-100. Over two years, a 100-item test was developed. The purpose of the first experiment was to demonstrate that the test would discriminate black from white takers. One hundred white and 100 black high school students ranging in age from 16 to 18 years, half from low socioeconomic levels and half from middle income levels, from the city of St. Louis took the BITCH-100. The black group showed a clear superiority over the white group. The distribution of scores approximated a normal distribution in which blacks comprise the upper half, whites the lower half. Twenty-eight black Neighborhood Youth Corps high school "drop outs" were administered the BITCH and the California Achievement Test in the second experiment. The results confirm the hypothesis regarding the sensitivity of the BITCH in picking up "intellectual indicators" not commonly found in conventional tests. (Author/JM)

A pre-publication release. Not to be quoted without the author's written permission.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

The BITCH-100: A Culture-Specific Test[1]

By Robert L. Williams, Ph.D.
Professor of Psychology
and Director of Black Studies
Washington University
St. Louis, Missouri

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Robert L. Williams
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

ED 070799

Developmental work on the BITCH (Black Intelligence Scale of Cultural Homogeneity) began approximately two years ago. The aim of the present paper is to describe the rationale and evaluation of the BITCH-100.

Approximately two decades ago, psychologists and educators devoted considerable effort toward devising intelligence tests whose items were equally fair to persons from various socio-economic levels. Especially during the period 1950 to 1960, many articles, pertaining to "culture-free" culture-fair" and "culture-common" tests, appeared in the literature. Although several tests claiming to be culturally fair were constructed during the 50's none proved of great significance. The findings generally showed lower predictive validity for culturally fair tests than for conventional ones. (Anastasia, 1968). In a recent report, however, Williams (1972) pointed to certain fallacies in statistical forecasting, particularly when the moderator variable is taken under consideration. In his presidential address to Division 5 at the 1967 annual APA meeting, Wesman (1968) concluded that the search for a culture-fair test was "sheer nonsense". This writer essentially agrees with Wesman but for different reasons. Since the American society is pluralistic on the one hand and racist on the other, it would be virtually impossible to conceptualize an instrument which would be fair to all people: Asians, Blacks, Caucasians, Chicanos, Indians, and Puerto Ricans.

UD 013110

[1]A paper presented at the American Psychological Association, Honolulu, Hawaii, September, 1972. This research was supported in part by grant #21557 from the National Institute of Mental Health. This paper will be published in the author's forthcoming book, Contemporary Issues in Black Psychology.

Although the search for culture-fair tests has been intelligently criticized, an equally strong objection can be raised against norm-referenced and other conventional tests. In light of the methodological and theoretical difficulties involved in developing culturally fair and culturally free tests, it is necessary therefore to examine several alternative considerations in test construction.

In addition to culture fair and culture free tests approximately five other approaches to test construction have been described in the literature:

1) Norm-referenced tests
2) Criterian-referenced tests
3) Learning potential assessment devices
4) Dialect-fair tests
5) Cultural specific tests

I. Norm-referenced tests:

Norm-referenced measures are by far the most popular of all methods used. A norm-referenced test is basically a standardized measure which has been administered with standard directions under standard conditions to a sample of examinees who are supposedly representative of the group for whom the test is intended. The purposes of standardization procedure are to obtain 1) a set of scores which will yield a normal distribution and 2) a set of norms for such factors as age, sex and grade. For a distribution of test scores to be "normal", one-half of the group must have scores above the mean whereas the other half must have scores below the mean. Any items which do not contribute to the normal distribution are discarded. Most test manuals contain tables of norms so that a person's standing may be determined by comparing his raw score to that of the reference group.

For many years the biases of the Stanford-Binet, the Weschler Intelligence Scale for Children, and the Peabody Picture Vocabulary Test, have been well known and well publicized. In fact, Weschler (1944) Clearly warned that his Weschler Bellevue test norms were to be used exclusively for the white population:

> "We have eliminated the colored versus white factor by admitting at the outset, that our norms cannot be used for the colored population of the United States. Though we have tested a large number of colored persons, our standardization is based upon white subjects only. We ommitted the colored population from our first standardization because we did not feel that norms derived by mixing the populations could be interpeted without special provisos and reservations (Pg. 107)."

In addition, three of the most prestigious individual ability tests, the Binet, Weschler, and Peabody, sysmetically excluded Black children from the normative samples. The 1937 Stanford-Binet, standardized on 3,184 American-born white children, was in use 23 years before being replaced by the 1960 form LM revision. The latter used, 4,498 subjects in the normative sample. The WISC was standardized on a sample of 2,200 white boys and girls (Weschler 1949). Another popular intelligence test, the Peabody Picture Vocabulary, excluded Black children from its standardization sample; 4,012 white children were used in the sample. Thus, no Black children were included in several of the major individual tests for children. Norm-referenced tests have been exclusive and non-representative rather than inclusive and representative.

II. Criterion Referenced Measures

Criterion referenced measures may prove to be a strong breakthrough in testing. What are criterion-referenced measures? According to Glaser (1963) they are measures which depend on an "absolute standard" as opposed to norm-referenced measures which depend on "relative standard". Thus, the basic difference between norm-referenced and criterion-referenced is the standard against which a student's performance is judged. Livingston (1972) states:

> When we use norm-referenced measures, we want to know how far a student's score deviates from the group mean. When we use criterion-referenced measures, we want to know how far his score deviates from a fixed standard, the criterion.

Items on criterion-referenced measures represent a sample of tasks which were drawn from a universe of instructionally relevant tasks. Each test item is selected solely on the basis of its content validity without regard to its discriminatory ability. In norm-reference tests, however items which do not discriminate are rejected and thrown out. The objective of criterion-referenced

to get the student to perform at a particular minimum acceptable level before he is permitted to go on to a higher level.

A student does not really fail the entire criterion-referenced test. He may be credited with previous achievements. The student is also permitted to proceed at his own rate of mastery rather than rank comparison with normative group children. Some criterion-referenced tests use 70 percent mastery; others 80 percent; still others as high as 90 percent. Although not a panacea this approach tends to provide the kind of information needed to help students in the educational process rather than label and mislabel their efforts.

III. Learning Potential Assessment Devices

Traditionally the term aptitude referred to a person's ability to profit from further training, or to acquire new knowledge and proficiency with training. Aptitude was considered to be a combination of in-born ability and acquired skills. Thus, an aptitude test was designed to measure the potential ability or capacity of a person to learn various skills. In a sense, the term aptitude is a mis-nomer since basically it is an achievement test. If the aptitude test is redefined to mean a test that predicts future accomplishment, then achievement tests, since they may be used to predict future accomplishment, are also aptitude tests.

In order to clear the confusion, an alternative approach has been introduced, called Learning Potential Assessment Devices, in which the student's rate of learning is assessed. Budhoff (1969) has developed a technique which shows the extent to which a stubject is a gainer or a non-gainer after he has been coached. Budhoff has identified three basic groups of learners: (1) high scorers or those who do well on both pre-test and past-test (2) subjects who perform poorly on the pre-test, but markedly improve their scores following coaching (gainers) and (3) others who perform poorly on the initial trial and fail to demonstrate improvement following training (non-gainers). The learning potential concept is process oriented and is derived from a conception in which

intelligence is defined as the ability to profit from problem relevant experience. The focus is on the child's educability and the trainability of cognitive processes. The learning potential measurement paradigm replaces the one-shot testing model with a three-stage program: (1) pre-test, (2) coach, (3) post-test. The pre-test allows the subjects to familiarize themselves with the demands of the task. The coaching session, which immediately follows, provides relevant problem solving strategies for the reasoning task. The post-test score includes both the child's initial ability and the effects of his learning. Potentially able but culturally different, children may thus be expected to show substantial improvement from pre to post-test. (Babad and Budhoff, 1971) Black children do not have the same experiences which facilitate the spontaneous acquisition of school-relevant skills, and tend to perform poorly on I.Q. tests. Their low I.Q.'s reflect cultural differences rather than inferior mental capacities. The learning potential paradign minimizes the effects of these cultural differences by providing all subjects with appropriate experiences relevant to dealing with the task. Differences and abilities among subjects have been reflected in their level of competence following appropriate training.

IV. Dialect-Fair Tests

With the current interest in urban education and the language of the culturally different Black student, educators have been looking for new methods that might prove useful in teaching standard English. Also, efforts have been made to reconstruct the language of tests in dialect that is fair to the Black child. Williams and Rivers, (1972b) showed quite clearly that test instructions in standard English penalized the Black child. If the language of the test is put in familiar labels without training or coaching, the child's performances on the test increases significantly. Little consideration has been thus far

given to the problems which dialect differences pose in test construction. Cadzen (1966) has already pointed out the following:

> "Ideally a child's language development should be evaluated in terms of his progress toward the norms for his particular speech community. This issue of "dialect fair" scales of language development may become as significant in thefuture as that of "culture fair" tests of intellignce has been in the past."

Many tests penalize the child in the usage of language as pointed out by Williams and Rivers (1972b).

V. A Cultural Specific Approach.

In spite of the many efforts made to develop culturally fair and culturally free tests, none has been developed. Williams (1970) suggested constructing a test based on items drawn exclusively from the Black culture. In an eloquent presentation, Barnes (1972) also suggested the cultural specific method:

> "Perhaps a potentially more fruitful approach lies in the development of "culture specific tests". If this suggestion seems far out, then ponder this. The model for culture-specific tests already exists, and when appropriately used, displays considerable effectiveness. Consider for example, the Stanford-Binet, and the WISC. These are examples of "culture specific tests". The culture in this instance is what is frequently referred to as "white middle class"....The point is that "culture specific" tests could be used to determine the child's ability to function symbolically or to think in terms of his own culture and environment. After all, this is what the S-B does for the white child. If a child can learn in one environment he can learn in another. If a child from the Mississippi Delta has learned the relationship between "Red Bone" and "Blue Tick" or between "Sweet Milk" and "Poke Salad", or whether to run from or cook a "Tedder", that child demonstrates the same capability for conceptual thinking as the middle-class white child, who has learned the relationship between "piano" and "violin". If he can learn these relationships in his own culture, he can also master those aspects of the elementary school curriculum, requiring this dimension of ability". (pg. 6)

Aside from the Bitch test, several culture specific test have appeared on the scene in the past few years. One of the earliest and highly popularized test was the Dove Counterbalance Intelligence Test. Another effort was made by Howard Lyman and students at the University of Cincinnati in 1970. First

called the Checkerboard Test and now the American Cross Cultural Ethnic Momenclature Test, ACCENT (Form A). The instrument contains 20 Black biased and 20 white biased items. It was administered to 110 undergraduates (91 whites and 19 Blacks) in education. The Black students obtained a mean of 15.3 on the Black items and 11.1 on the white items, a difference found to be significant at the point .001 level of confidence. White students obtained a mean of 12.7 on the white items and only 8.3 on the Black items, a difference significant also at the point of .001 level of confidence.

Another instrument was developed by junior high school students from Des Moines, Iowa High School. This test is reported in the NEA Center for Human Relations, 1972. This scale is called the S.H.A.F.T. or Students Hype, Arranged for Teachers. The test is mainly used for sensitivity and awareness sessions. At the 10th National Conference on Violations of Human and Civil Rights: Test and Use of Tests by the National Education Association, the Shaft test was administered to 650 conferees. A few scored high, many were average and a few "bummed it".

Culture specific tests have the advantage of dealing with content material which is familiar to the Black child. This means that he already has stored away mental images of the material so that he does not have to deal with the foreign or unfamiliar aspects of these materials. Thus, a combination of dialect specific and culture specific tests would certainly enhance the possibility of measuring accurately what is inside the Black child's head. This is the basic rationale for the BITCH-100.

Experiment I

I. Method of Procedure

a) Selection of test items

The BITCH-100 is a culture-specific test. It is not intended to be a culture-fair or culture-common test. The research has been guided by several considerations. The first problem facing anyone attempting to devise a test is the selection of items. For the BITCH, the content of all items was drawn exclusively from the Black experience domain. A fairly comprehensive selection of items was made from a variety of sources. The words included were taken from the Dictionary of Afro-American Slang, the Word in the APGA Journal, friends and the author's personal experiences gained from living and working in the Black community. The original word list consisted of 175 randomized items.

The second consideration of the BITCH-100 developmental work was item objectivity. All items were edited to eliminate careless phraseology, ambiguity and duplication. The more objective an item is, the higher the test reliability. Thirdly, the items were administered to Black and white experimental groups, in order to identify 1) criteria for defining words 2) items common to Black and white groups and 3) words which pulled associations peculiar to white groups. By this method, the words which seemed to discriminate poorly between whites and Blacks were quickly eliminated. The fourth step involved tryout sessions with a group of four judges, two Black and two white, who rated the items for ambiguity, clarity and objectivity. The tryout sessions proved extremely helpful in dulling some items and sharpening others. Final item selection consisted of the best 100 from the original 175 items. (See Figure I)

b) Subjects

The subjects were chosen from the city of St. Louis. 100 white and 100 Black Ss were used in Experiment I. All Ss were high school students ranging in age from 16 to 18 years. Half of the Ss were from low socio-economic levels whereas the other half came from middle income levels.

c) Instructions

The administration of the BITCH-100 is fairly simple. The test requires less than one-half hour to administer. The following directions were given:

> Directions: Below are some words, terms and expressions taken from the Black experience. Select the correct answers and put a check ( ) mark in the space provided on the right of the test sheet. Remember, we want the correct definition as Black people use the words and expressions. There is no time limit. 20 to 30 minutes should be sufficient to complete the test. Go ahead.

Experience with the tryouts during standardization indicated test virtually all Black subjects became intensely interested in the test. Comments were made such as : "Man, this is a bad test." "This is really hip." "It's outta sight." Black Ss frequently came across items which were humorous and quite familiar to them. White Ss seemed to be quite challenged by the test and appeared tense. Many sighed and showed other signs of discomfort. A few questioned the validity of the instrument; others stated that if the test is valid, then they have little knowledge of the Black experience.

Results and Discussion

The means, standard deviations of the BITCH scores of the Black and white groups separately and combined are presented in Table I. The Black group shows a clear superiority over the white group of 36.00 mean points, a difference that is significant at the level of confidence. It is also clear that the shape of the two distributions of the white and Black groups are different. Both curves are asymmetrical and deviate significantly from the Bell-shaped distribution. The usual rationale for skewness is defined in terms of the "difficulty" of items. On the one hand if the items of a test are rather easy, high total scores will be in greater abundance than low

scores and the distribution will be negatively skewed reflecting an elongated tail on the low side of the curve (DuBois, 1965). Such is the shape of the distribution for the Black group. The traditional argument is straight forward and simple; the items were easy for the Black Ss. Another more contemporary interpretation might be that the students all have average to above average ability and that this ability is tapped by the BITCH-100.

On the other hand, if the test is composed of very difficult items, low scores will predominate, high scores will be relatively rare and the distribution will show positive skewness; that is the tail at the high end of the curve will be elongated. Such is the case of the shape of the distribution for the white group. The items were "difficult" for whites and "easy" for Blacks.

The basis question here is one dealing with whether the test items created the distributions obtained or if the underlying trait (abilities) of individuals caused the distributions to be both skewed and bi-modal. The interpreation considered here is twofold: 1) a culture specific test clearly shows the abilities of the group for which the test was intended and 2) a culture specific test does not accurately reflect the abilities for a non-representative group.

Table II provides norms derived from T-scores white and Black groups separately and combined. Certain details in Table I are noteworthy. First, the separate norming process for the white and Black groups indicates the extent to which a test that is normed on one group and used on another is patently unfair. The BITCH-100 is designed primarily for the Black experience. Whites are clearly penalized. Using the Black norm as a basis for determing the value of a white student's score, it is clear that most white Ss would

generally score at the lower end of the distribution. They would need a score of 87 in order to earn a T score of 50 or at the mean of the distribution. Moving then from the Black norm to the combined Black and white norm a score between 68 and 70 is needed to place at the mean. Inspection of the raw data shows that only 8 out of the 100 white Ss scored 67 or better, whereas all but one Black subject earned scores above 80. What is important is that the combined distribution approximates the so-called normal distribution in which half of the scores fall above the mean and half fall below the mean.

Traditionally only a few Black students were in the range above the mean. Black and lower SES children comprised the lower half of the "normal" distribution. In the present study, however, the situation is reversed. Blacks comprise the upper half of the distribution, whereas whites comprise the lower half. The other important point is the severe penalty by whites on a culture specific test, just as severe as Blacks experience on other culture specific tests as the Binet, WISC and Peabody.

A further point to be made is that a normal curve is a theoretical curve; it is an assumption, based on probability theory. As mentioned earlier so-called easy test items, as a rule, do not discriminate. They are therefore discarded. In the context of this paper, the definition of easiness or difficulty is relative. What is easy for one group is difficult for another group. Proponents of testing claim that a negatively skewed test is useless because the items do not measure individual differences among the more able subjects of the group. They suggest including more difficult items in order to insure a normal distribution. This writer would have to question this search for individual differences particularly in a society that is ostensibly based on equality. Anastasia (1968) claims that a test is "excess baggage" if everyone

passes every item. If everyone passes every item, one might have a great deal of information about the group as a whole. In disagreement with Anastasia, such a test might prove more reliable than one whose items are scaled according to difficulty. The test may be more valid than one which reflects individual differences. Suppose for example, on a job selection test of 100 applicants 50 percent fail the test and 50 percent pass. Does this mean that the ones who passed the test will perform significantly better on the job than those who failed? Not so in all instances. Coupland (1970) cites the following situation of test discrimination that resulted in employment discrimination:

> In one of the local plants studies, fifteen Negro workers were employed on a production line in assembly work. Due to production pressures, these workers were hired without the usual Wunderlicht battery of tests. After a six-month period, the workers were given the tests. In spite of the fact that each of the workers had received a satisfactory supervisor rating on the job, not one of the fifteen received a passing test score! (p. 244)

Such a finding is no accident. It is not a unique incident. A similar serendipitous finding was noted in the hiring of minority postal employees in San Francisco. They were not given the usual screening test at the beginning of employment. However at the end of one year, they were all tested and received failing scores. Yet virtually all employees had received satisfactory supervisory ratings for throwing mail. It would appear that the best criterion for throwing mail is throwing mail rather than a score on a test. Thus, the validity and reliability of a test may be substantially increased if the test yield a graater than 50 percent pass rate. However, one would have to alter one's assumption about the distribution of the trait under consideration. For example, throwing mail may not be normally distributed. Also, the trait measured by the BITCH-100 may not be normally distributed in the Black population.

Experiment II

Next, we come to the knotty problem of validation. How do we know that the BITCH is measuring intelligence rather than some other phenomenon? My honest reply is that I do not know. Only practical experiences will validate the BITCH.

The usual procedure for validating a new test, however, is to set as a criterion a well-established test (e.g. the Binet or the WISC) which has been accepted as a "good" measure of intelligence. The next step is to determine the relationship between the old and new tests. The significance of this correlation will depend of course upon the validity of the original criterion. Thus, it is frequently the criterion per se rather than the new test, which needs examination. The general tendency has been to accept tests already in existence as established measures of Black intelligence. This is not the case. The final validation of the BITCH will not rest on how well it correlates with established ability tests, but how well it works out in practice.

Since culture-specific tests are considered fair for particular groups they must be validated on a criterion that is not a conventional test. Such validation is especially necessary if the generally accepted criterion is not valid for Black children or as Williams (1972) showed that the predictor is just as biased as the criterion. There is a need, however for a culture specific test just as there existed a need for the WISC which was developed because of the belief that the Binet was deficient.

If new test is markedly out of step with conventional measures, the question is whether it can serve as a valid and reliable measures of intelligence. Again, the degree to which the BITCH correlates with conventional tests cannot be accepted as a basic test of the BITCH's validity. The test must stand on its own two feet. To test out the extent to which the BITCH

correlates with a conventional measure, a special study was undertaken to determine the relationship between the BITCH and the California Achievement Tests.

Procedure:

Twenty-eight Black Neighborhood Youth Core high school "drop outs" were administered the BITCH and the California Achievement Test (CAT. The Ss were 17 females and 11 males ranging in age from 16-18 with a mean age of 17.8. The sequence for test administration was the CAT on the first day and the BITCH on the second day.

Results and Discussion

The coefficients of correlation between the three achievement subtest scores of the CAT and teh BITCH, along with the means and standard deviations, are presented in Table III. The obtained grade levels are substantially below what might be expected on the basis of the age levels. Assuming that 17 year olds, by the usual standards, should be completing the 11th grade or entering the 12th the mean grade levels or the CAT are quite below expectation. On the BITCH, however, a mean of 80.79 yields a T-score of 55 indicating the extent to which a fair test leads to greater comprehension and better performance on the instrument. The results confirm the hypothesis regarding the sensitivity of the BITCH in picking up "intellectual indicators" not commonly found in conventional tests.

As might be expected, there is a low correlation between the three subtests of the CAT and the BITCH. Thus, the Ss who scored low on the CAT did not necessarily score low on the BITCH. To the contrary, some of the low CAT scorers were among the high BITCH scorers. The relatively high CAT scorers were not necessarily the top BITCH scorers. These findings suggest that the BITCH and the CAT may be measuring different phenomena.

Conclusions

Obviously the BITCH-100 as a culture specific approach represents a different approach to psychological testing. These early developments however, indicate that something is wrong with the present testing system which places such great emphasis on individual differences.

The BITCH-100 can be used in ways other than test of a cognitive funcation. For example one dissertation is under way currently which utilizes the BITCH-100 as a predictor of empathy in whites who counsel Blacks. Other uses may be with measuring awareness and familiarity of white with the Black experience. Attempts can be made to examine change scorers or the extent whites are willing to engage themselves in the Black experience.

Additional research is needed on the BITCH-100. Currently a sample of approximately 54,000 Black students in four regions of the country are being tested. A split-half reliability study will be conducted to determine test consistency. In addition regional variations will be examined. Also, several studies are under way which involve further validation of the BITCH-100 using non-traditional criteria.

TABLE I

Means and Standard Deviations of Blacks and Whites on the BITCH-100

	White	Black	Combined W and B
N	100	100	200
Mean	51.07	87.07	69.07
Sd	16.20	6.97	21.92

TABLE II

Profile Sheet Black, White and Combined BITCH-100 Scores

BITCH RAW SCORES

T-SCORES	WHITE	BLACK	TOTAL B&W
80	99-100		80
75	91-92		75
70	84	100	70
65	75-76	97	99-100 65
60	67-68	94	90-93 60
55	59	91	79-81 55
50	51	87	68-70 50
45	43	84	58-59 45
40	35	80	47-48 40
35	26-27		36-37 35
30	18	73	25-26 30
25	10		14-15 25
20	2-3	66	3-4 20
10		59	10
5		56	5
	White W	Black	Total B&W

TABLE III

Coefficients of Correlation Between
CAT and BITCH Scores

Test	n	Mean	Sd	r with BITCH
Reading	28	7.60	1.96	.39
Language	28	7.69	1.81	.33
Math	28	7.34	1.34	.18
BITCH	28	80.79	9.20	-

REFERENCES

Anastasia, A., Psychological Testing, (Third Edition), Toronto, Ontario: McMillan Company, 1968.

Anonymous. The word. Personnel and Guidance Journal, 1970, 48, p. 69.

Babad, E., and Budhoff, M., Sensitivity of Learning Potential Measurement in Three Levels of Ability, Studies in Learning Potential. Research Institute for Educational Problems, Cambridge, Mass., 1971, pp. 1-14.

Barnes, E., I.Q. Testing and Minority Children: Imperatives for Change, 1972. National Leadership Institute Teacher Education/Early Childhood. The University of Connecticut, Technical Paper, pp. 1-8.

Budhoff, M., Learning Potential: A Supplementary Procedure for Assessing the Ability to Reason. Seminars in Psychiatry, 1969, 1, pp. 278-290.

Cadzen, C. B., Subcultural Differences in Child Language: An Inter-disciplinary Review. Merrill-Palmer Quarterly, 1966, 12, pp. 185-214.

Callahan Jr. High School Students, Des Moines, Iowa. The S.H.A.F.T. Test, Report o of the Tenth National Conference on Civil and Human Rights in Education, Violations of Human and Civil Rights: Tests and Use of Tests, National Education Association, Washington, D.C.

Coupland, D. E., Aptitude Tests and Discrimination, International Labour Review, 1970, 102, pp. 241-253.

*Glaser, R., Instructional Technology and the Measurement of Learning Outcome, American Psychologist, 1963, 18, pp. 519-521.

Livingston, S., Criterion-referenced Applications of Classical Test Theory, Journal of Educational Measurements, 1972, 1, pp. 13-26.

Major, C., Dictionary of Afro-American Slang. New York: International Publishers.

Wechller, D., The Measurement of Intelligence. Baltimore: Williams and Williams Co., 1944.

Wechsler, D., Manual for the Wechsler Intelligence Scale for Children. New York: The Psychological Corporation, 1949.

Wesman, A.G., Intelligence Testing, American Psychologist, 1968, 23, pp. 267-274.

Williams, R.L., Danger: Testing and Dehumanizing Black Children, Clinical Child Psychology Newsletter, 1970, 9, pp. 5-6.

Williams, R.L., The Problem of Match and Mis-Match in Testing Black Children, a paper read at the Annual Meeting of American Psychological Association, Honolulu, Hawaii, 1972. (a)

Williams, R.L., and Rivers, L.W., Cognitive Development in Black Children: Non-Standard English or Different Strokes. An unpublished manuscript, Washington University, Black Studies Program, St. Louis, Mo., 1972.

Williams, R. L., and Rivers, L.W., The Use of Standard and Non-Standard English in Testing Black Children. A paper presented at the Annual Meeting of the American Psychological Association, Honolulu, Hawaii, September, 1972. (b)

*Dubois, P. H., An Introduction of Psychological Statistics. New York: Harper and Row Publishers, 1965.